Comparative functional genomics approach for the annotation of proteins in Unclassified Halophilic archaeon DL31

N. S. Sardeshmukh, S. G. Sanmukh* and E. Nakkeeran
Vellore Institute of Technology (VIT) University, Vellore Campus,
Vellore - 632 014, Tamil Nadu, India

(*Corresponding author's E-mail.: swamukh1985in@rediffmail.com)

Abstract— The structure, function and sub-cellular location prediction for the unknown proteins from Unclassified Halophilic archaeon DL31 were carried out for characterization of the proteins in their respective families. The 991 genes for hypothetical proteins in Halophilic archaeon DL31 chromosome were predicted by the application of computational methods and Bioinformatics web tools. The structure predictions for 206 unknown proteins were possible whereas functions were predicted in 825 protein sequences. The function prediction for the proteins were done by using Bioinformatics web tools like CDD-BLAST, INTERPROSCAN and PFAM by searching protein databases for the presence of conserved domains. The Sub-cellular location predictions were done for all the unknown proteins by using CELLO v 2.5 server. While tertiary structures were constructed using PS^2 Server- Protein Structure Prediction server. This study revealed structural, functional and Sub-cellular localization of unknown proteins in Unclassified Halophilic archaeon DL31chromosome.

Keywords— sub-cellular location; Bioinformatics web tools, conserved domains, Protein structure prediction.

I. INTRODUCTION:

Studies based on culturing showed a decrease in prokaryotic diversity recovered on plates as salt concentration increased (Rodríguez-Valera *et al*., 1981; 1985; 1988).At salinity up to 15%, most isolates were those, which are commonly found in seawater. Above 15% salt concentration, most strains are specialized halophilic. However, when a clone library from a crystallizer pond (salinity around NaCl saturation) was carried out, the most frequent 16S rDNA archaeal sequence (the SPhT phylotype) did not correspond to any previously known microorganism (Benlloch *et al*., 1995a). The Halobacteriaceae form a monophyletic group within the phylum Euryarchaea of the domain Archaea. The Halobacteriaceae include 26 named genera (each with at least one cultured species -see http://www.the-icsp.org/taxa/halobacterlist.htm (Oren, 2008))Halobacterium (abbreviated as Hbt.), Haloadaptus, Halalkalicoccus, Haloarcula (Har.), Halobaculum, Halobiforma, Halococcus, Haloferax (Hfx.), Halogeometricum, Halomicrobium, Halopiger, Haloplanus, Haloquadratum (Hqr.), Halorhabdus, Halorubrum, Halosimplex, Halostagnicola, Haloterrigena, Halovivax, Natrialba, Natrinema, Natronobacterium, Natronococcus, Natronolimnobius, Natronomonas (Nmn.), and Natronorubrum. As with most other groups of bacteria and archaea, many lineages are only known through rRNA-based studies of uncultured organisms (Sorensen et al., 2005). It has long been recognized that *Archaea* originating from a variety of diverse environments are able to N glycosylate numerous proteins (Eichler and Adams, 2005). Still, very little is known

about certain Asn residues in archaeal glycoproteins and the pathway responsible for the covalent attachment of glycan moieties. In comparison to other groups of extremophilic microorganisms such as the thermophiles and the alkaliphiles, the halophiles of all three domains have been relatively little exploited in biotechnological processes, with notable exceptions of β-carotene from *Dunaliella*, bacteriorhodopsin from *Halobacterium*, and ectoine from *Halomonas* (Oren, 2010). In addition, attention was drawn toward secondary metabolites from halophiles as well as bioremediation and biofuel production. Ectoine is the active ingredient of many cosmetics and skin care products and increasingly gaining importance in medicinal preparations (Graf et al., 2008). In addition, ectoine (and/or suitable derivatives) is used as a protectant for biomolecules and enhancer in molecular biology applications such as PCR and DNA microarray techniques (Mascellani et al., 2007, Schnoor et al., 2004). Production of poly-β-hydroxyalkanoates biodegradable polymers with plastic-like properties—although not restricted to halophilic prokaryotes. Some halophilic or halo tolerant *Bacteria* were shown to be excellent producers of such Bioplastic. One of these is *Halomonas boliviensis*, as argued by Jorge Quillaguama´n (Cochabamba, Bolivia), who presented strategies to optimize the biosynthesis of such bioplastics coupled with production of the high-value products ectoine and hydroxyectoine (Quillaguaman et al., 2010, Van-Thuocet al., 2010). *Archaea* of the genus *Haloferax* are also known as poly-β-hydroxyalkanoates producers and Hua Xiang and colleagues (Beijing, China) (Han et al., 2009, Lu et al., 2008) elucidated the biosynthetic pathways leading to their production.

Many alkaliphiles are halophilic as well, and many useful enzymes applied in the detergent industry (washing powders), the textile industry, and other processes were derived from bacteria growing in saline alkaline lakes and are already explored for the production of commercially valuable enzymes, in particular, proteases and amylases. Halophilic enzymes (typical for *Archaea* and *Salinibacter* but also for exoenzymes of any halophile) are characterized by an excess of acidic amino acids and subsequent negative surface charge. This peculiarity allows effective competition for hydration water and enables function in solutions of low water activity (viz. organic solvent/water mixtures.).

The worldwide problem of petroleum contamination and potential application of halophiles for bioremediation can be neutralized by a novel isolate, similar to *Alcanivorax dieselolei*, able to grow on crude oil, diesel fuel, and pure aliphatic hydrocarbons but unable to degrade aromatic compounds. Culturable strains of *Marinobacter* and *Halomonas* sp. was able to degrade various polyaromatic hydrocarbons over a salinity range from 1 to 17% NaCl and will become increasingly important in the future for their degrading potential of halophiles. Halophilic/haloalkaliphilic and halotolerant bacteria could be used to break down biomass material and form biofuel products. The required alkaline pretreatment (to remove lignin) and subsequent partial neutralization will create an environment for halophilic or haloalkaliphilic fermentative bacteria in cellulose-converting processes. The general trend toward use of algae for biofuel (biodiesel) production is problematic because of the high consumption of fresh water. It is thus possible that in the future the biotechnological application of halophiles, or of genes derived from them, will extend too many more members of this extremely diverse group of microbes. The presence of PHA granules in haloarchaeon was first reported in 1972 (Kirk and Ginzburg 1972). The strains were called at that time "Halobacterium sp. from the Dead Sea", but later

identified as Haloarcula marismortui (Oren et al. 1990). Since then, strains of several other haloarchaeal genera, including Haloferax, Halobiforma, and Haloquadratum, have been found to accumulate PHAs, such as poly (3-hydroxybutyrate) or poly (3-hydroxybutyrate-co-hydroxyvalerate) (Fernandez-Castillo et al. 1986; Hezayen et al. 2002a; Burns et al. 2000). PHA production by halophilic Archaea and Bacteria, with a focus on Haloferax mediterranei, which shows the highest potential of an archaeal source for industrial applications, and the characterization of enzymes involved in synthesis of PHA. The present paper reports the functional properties along with their structural and sub-cellular localization of unknown proteins present in chromosome of Halophilic archaeon DL31which will prove to be helpful for identifying novel enzymes and protein candidates with possible applications in the near future.

II. MATERIALS AND METHODS

Sequence Retrieval

The Complete sequence of chromosome of Halophilic archaeon DL31 was retrieved from the KEGG database (http://www.genome.jp/kegg/) (Lucas, et al., unpublished).

Functional Annotations

The screening of the sequenced genes for the presence of conserved domains using the web-tools were carried out by using bioinformatics web tools like CDD-BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) (Altschul et al., 1997; Schaffer et al., 2001; Aron et al., 2006),, INTERPROSCAN (http://www.ebi.ac.uk/Tools/pfa/iprscan/) (Zdobnov and Rolf., 2001; Quevillon et al., 2005), Pfam (http://pfam.sanger.ac.uk/) (Alex et al., 2004) and CATH (http://www.cathdb.info/) (Orengo et al., 1997)were used, which shows the ability to search the defined conserved domains in the sequences and assist in the classification of proteins in appropriate family (Sanmukh et al., 2011; Kanz,C. et al., 2005).

Functional Categorization

The function prediction web tools have shown variable results depending upon the information available in databases, when searched for the conserved domains or functional sites in the submitted protein sequences under study.

Sub-cellular Localization of the hypothetical proteins

The CELLO v 2.5 is a Sub-cellular Localization predictor server (http://cello.life.nctu.edu.tw/) which identifies the sub cellular location of given DNA and/or protein sequences of Gram positive, Gram negative and Eukaryotic cells. The identification of sub-cellular localization of the proteins is helpful in better understanding

of their functional properties that in turn are helpful in proper classification of unclassified proteins in their respective families (Yu et al., 2004; Yu et al., 2006).

Protein Structure Prediction

The 3D-structures of isolated protein genes were generated using online PS^2 Protein Structure Prediction Server (http://ps2.life.nctu.edu.tw/). The server accepts the protein (query) sequences in FASTA format and uses the strategies of Pair-wise and multiple alignments to generate resultant proteins 3D structures, which are constructed using structural positioning information of atomic coordinates for known template in PDB format using best scored alignment data. Where the selection of template was based on the same conserved domain detected in the functional annotations and which must be available in the structure alignment for modeling purpose.

III. RESULTS AND DISCUSSION

The comparative genomic studies for characterizing 991 genes of Unclassified Halophilic archaeon DL31 was done. The classifications of proteins were done by using sequence similarity search with close orthologous family members available in various protein databases using the web tools. The sub-cellular localization of the proteins was also predicted for assigning proper functional character/s. The online-automated PS^2 server was used for the prediction of 3-D structure of protein. The analysis of proteins by using web tools for classification of 825 proteins into particular protein family based on conserved domain available in the sequence are represented in respective Table 1. The $(PS)^2$ Server fabricate the three dimensional structures for 206 proteins satisfactorily using best scored orthologous template which are represented in Table 1. The templates having best scoring with protein sequences are represented in the order as Template ID, Identity, Score and E-value which represented in structure column.

IV. CONCLUSION

These in-Silico studies for characterization of unknown proteins of Halophilic archaeon DL31 were carried out for verifying the authenticity of the sequenced gene products. Bioinformatics Web Tools have shown the ability to predict structure and functions in such proteins successfully. We were able to predict and categorized 825 proteins functionally and 206 proteins structurally from 991 hypothetical protein sequences screened from the chromosome of unclassified Halophilic archaeon.

ACKNOWLEDGMENT

One of the authors Mr. Nikhil S. Sardeshmukh wants to thank Dr. E. Nakkeeran and Mr. S. G Sanmukh for providing support and assistance in for completing this work successfully.

REFERENCES

1. Alex, B., Lachlan, C., Richard, D., Robert, D. F., Volker, H., Sam, G.J., Ajay, K., Mhairi, M., Simon, M., Erik, L. L. S., David, J. S., Corin Y., Sean, R. E., (2004). The Pfam families' database. Nucleic Acids Research, Vol. 32, D138-D141.
2. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J., (1997). Gapped BLAST and PSI-BLAST: "a new generation of protein database search programs". Nucleic Acids Res. 25 (17), 3389-402.
3. Aron, M. Bauer., John, B. A., Myra, K. D., Carol, D. S., Noreen, R. G., Marc, G., Luning, H., Siqian, H., David, I. H., John, D. J., Zhaoxi, K., Dmitri, K., Christopher, J. L.,Cynthia A. L., Chunlei, L., Fu, L., Shennan, L., Gabriele, H. M., Mikhail, M., James, S. S., Narmada, T., Roxanne, A. Y., Jodie, J. Y., Dachuan, Z., Stephen, H. B., (2006). CDD: "*a conserved domain* database for interactive domain family analysis. "Nucleic Acids Research, Vol. 35, D237–D240.
4. Benlloch, S., Martínez-Murcia, A.J., and Rodríguez-Valera, F. (1995a) Sequencing of bacterial and archaeal 16S rRNA genes directly amplified from a hypersaline environment. Syst Appl Microbiol 18: 574–581.
5. Burns DG, Camakaris HM, Janssen PH, Dyall-Smith ML (2000) Cultivation of Walsby's sqare haloarchaeon. FEMS Microbiol Lett 238:469–473
6. Eichler, J., and M. W. Adams. 2005. Posttranslational protein modification in archaea. Microbiol. Mol. Biol. Rev. 69:393–425.
7. Fernandez-Castillo R, Rodriguez-Valera F, Gonzalez-Ramos J, Ruiz- Berraquero F (1986) Accumulation of poly(-hydroxybutyrate) by halobacteria. Appl Environ Microbiol 51:214–216
8. Graf, R., S. Anzali, J. Bu¨nger, F. Pflu¨cker, and H. Driller. 2008. The multifunctional role of ectoine as a natural cell protectant. Clinics Dermatol. 26:326–333.
9. Han, J., Q. Lu, L. Zhou, H. Liu, and H. Xiang. 2009. Identification of the polyhydroxyalkanoate (PHA)-specific acetoacetyl coenzyme A reductase among multiple FabG paralogs in Haloarcula hispanica and reconstruction of the PHA biosynthetic pathway in Haloferax volcanii. Appl. Environ. Microbiol. 75:6168–6175.
10. Hezayen FF, Tindall BJ, Steinbüchel A, Rehm BHA (2002a) Characterization of a novel Halophilic archaeon, Halobiforma haloterrestris gen. nov., sp. nov., and transfer of Natronobacterium nitratireducens to Halobiforma nitratireducens comb. nov. Int J Syst Evol Microbiol 52:2271–2280
11. Kanz, C. et al. (2005) The EMBL Nucleotide Sequence Database. Nucleic Acids Res., 33, D29–D33.
12. Kirk RG, Ginzburg M (1972) Ultrastructure of two species of halobacterium. J Ultrastruct Res 41:80–94
13. Lu, Q., J. Han, L. Zhou, J. Zhou, and H. Xiang. 2008. Genetic and biochemical characterization of the poly (3-hydroxybutyrate-co-3-hydroxyvalerate) synthase in Haloferax mediterranei. J. Bacteriol. 190:4173–4180.
14. Lucas,S., Han,J., Lapidus,A., Cheng,J.F., Goodwin,L., Pitluck,S., Peters,L., Mikhailova,N., Teshima,H., Detter,J.C., Han,C., Tapia,R., Land,M., Hauser,L.,

Kyrpides,N., Ivanova,N., Pagani,I., Dyall-Smith,M., Cavicchioli,R. and Woyke,T. Complete sequence of chromosome of Halophilic archaeon DL31 (Unpublished).

15. Mascellani, N., X. Liu, S. Rossi, J. Marchesini, D. Valentini, D. Arcelli, C. Taccioli, M. H. Citterich, C.-G. Liu, R. Evangelisti, G. Russo, J. M. Santos, C. M. Croce, and S. Volinia. 2007. Compatible solutes from hyperthermophiles improve the quality of DNA microarrays. BMC Biotechnol. 7:82.
16. Mescher, M. F., and J. L. Strominger. 1976. Purification and characterization of a prokaryotic glucoprotein from the cell envelope of Halobacterium salinarium. J. Biol. Chem. 251:2005–2014.
17. Oren A (2008) Correct names of taxa within the family Halobacteriaceae – May 2008.
18. Oren A, Ginzburg M, Ginzburg BZ, Hochstein LI, Volcani BE (1990) Haloarcula marismortui (Volcani) sp. nov., nom. rev., an extremely halophilic bacterium from the Dead Sea. Int J Syst Bacteriol 40:209–210
19. Oren, A. 2010. Industrial and environmental applications of halophilic microorganisms. Environ. Technol. 31:825–834.
20. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM (1997). CATH--a hierarchic classification of protein domain. Structures.15: 5(8):1093-108.
21. Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R., Lopez R. (2005) InterProScan: protein domains identifier. Nucleic Acids Res. 33 (Web Server issue):W116-W120
22. Quillaguaman, J., H. Guzman, D. Van-Thuoc, and R. Hatti-Kaul. 2010. Synthesis and production of polyhydroxyalkanoate by halophiles: current potential and future prospects. Appl. Microbiol. Biotechnol. 85:1687–1696.
23. Rodríguez-Valera, F. (1988) Characteristics and microbial ecology of hypersaline environments. In Halophilic Bacteria.
24. Rodríguez-Valera, F., Ruíz-Berraquero, F., and Ramos-Cormenzana, A. (1981) Characteristics of the heterotrophic bacterial populations in hypersaline environments of different salt concentrations. Microb Ecol 7: 235–243.
25. Rodríguez-Valera, F., Ventosa, A., Juez, G., and Imhoff, J.F. (1985) Variation of environmental features and microbial populations with salt concentrations in a multi-pond saltern. Microb Ecol 11: 107–115.
26. Schaffer, A. A., Aravind, L., Madden, T. L., Shavirin, S. Spouge, J. L., Wolf, Y. I., Koonin, E. V., Altschul, S. F., (2001*).” Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements”*. Nucleic Acids Res. 29(14), 2994-3005.
27. Schnoor, M., P. Voß, P. Cullen, T. Boking, H. J. Galla, E. A. Galinski, and S. Lorkowski. 2004. Characterization of the synthetic compatible solute homoectoine as a potent PCR enhancer. Biochem. Biophys. Res. Commun. 322:867–872.
28. Sorensen KB, Canfield DE, Teske AP, Oren A (2005) Community composition of a hypersaline endoevaporitic microbial mat. Appl Environ Microbiol 71: 7352–7365.
29. Swapnil G. Sanmukh, Waman N. Paunikar, Tarun K. Ghosh and Tapan Chakrabarti 2011. Structural & Functional Prediction of Hypothetical Proteins In

Bacteriophages Against Halophilic Bacteria - An In Silico Approach. Int J Pharm. Bio. Sci. Vol 2 (2), B61-B70
30. Van-Thuoc, D., H. Guzman, J. Quillaguama´n, and R. Hatti-Kaul. 2010. High productivity of ectoines by Halomonas boliviensis using a combined two-step fed-batch culture and milking process. J. Biotechnol. 147:46–51.
31. Yu CS, Chen YC, Lu CH, Hwang JK: Prediction of protein subcellular localization. Proteins: Structure, Function and Bioinformatics 2006, 64:643-651.
32. Yu CS, Lin CJ, Hwang JK: Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. Protein Science 2004, 13:1402-1406.
33. Zafer, A., Yucel, A., Mark, B. Protein secondary structure prediction for a single-sequence using hidden semi-Markov models, BMC Bioinformatics ,7, 178, 2006.
34. Zdobnov, E. M., Rolf, A. Interproscan- an integration platform for the signatures recognition methods in InterPro. Bioinformatics 17,847-848, 2001.

Mr. Nikhil Suhas Sardeshmukh, born on 17th September 1988 in Solapur, Dist- Solapur; Maharashtra, India. He is pursuing his Masters in Biotechnology from VIT University, Vellore, Tamil Nadu, India. His major fields of interest are Process Engineering, Environmental Engineering, and Bioinformatics. (e-mail:179nikhil@gmail.com).

Mr. Swapnil Ganesh Sanmukh, born on 07th November 1985 in Miraj, Dist-Sangli, Maharashtra, India. He had completed his Masters in Environmental Biotechnology from Shivaji University, Kolhapur, Maharashtra in 2009. He is pursuing his Ph.D in Environmental Biotechnology from University of Madras, Chennai-600113, Tamil Nadu, India. His major fields of interest are Microbiology, Molecular biology, Biochemistry and Bioinformatics. (E-mail: swamukh1985in@rediffmail.com).

Dr. E. Nakkeeran, born on 04th May 1980 in Tamil Nadu, India. He had completed his Masters in Biotechnology from VIT University, Vellore, Tamil Nadu in 2005. He also completed his Ph.D in Biotechnology from University of Mysore, Karnataka, India. His major fields of interest are Microbiology, Molecular biology and Bioinformatics. He is now working as a Senior Assistant Professor in School of Biosciences & Technology (SBST), VIT University, Vellore-440020, Tamil Nadu, India (e-mail: bionakkeeran@gmail.com).